Automatic Detection of Low-Quality Seismocardiogram Cycles Using the Outlier Approach

V. Zakeri[1], F. Khosrow-Khavar[2], K. Tavakolian[3]

[1]Heart Force Medical Inc., Vancouver, BC, Canada
[2]Simon Fraser University/Engineering Science, Burnaby, BC, Canada
[3]University of North Dakota/Electrical Engineering, ND, USA

Abstract— **In this study, an algorithm was developed to automatically detect the low-quality (LQ) cardiac cycles in seismocardiogram (SCG). The proposed algorithm extracts some features from the SCG signal, which are referred to as signal quality indices (SQIs), and computes the outlier points of each SQI. Our hypothesis was that the identified cycles (outliers) would include the LQ ones. To verify this hypothesis, the algorithm results were compared with the LQ cycles that were labeled manually by an expert in the field. The developed algorithm was tested on total 1697 cardiac cycles, and there was a great overlap between the computed outliers and the LQ cycles (84% of 248 LQ cycles were identified). The proposed algorithm is simple, efficient, and works in an unsupervised manner.**

Keywords— **SCG, Signal quality indices, Outlier removal, Low-quality cycle**

I. Introduction

Seismocardiogram (SCG) is derived by recording the accelerations created by the heart from the surface of the chest. There have been different research works correlating certain features of SCG to hemodynamic parameters such as systolic time intervals as well as to myocardial abnormalities and diseases [1]–[3].

A challenge to overcome in processing SCG is the beat to beat variations that at times severely affects the morphology of the signal [4], making it difficult to extract the pertinent cardiac parameters. Such changes of morphology can be created through regular breathing [5], the inherent abnormalities of the heart such as myocardial ischemia [6], and/or the corruptions induced by noise and artifact. On the other hand, there is a consensus that around 10 cardiac cycles are sufficient for reliably deriving the cardiac parameters such as systolic time intervals [7], [8].

In this paper we have proposed a new algorithm for detecting and removing the low-quality (LQ) SCG cycles. The remaining cycles (at least 10) would have a higher quality, and can be used to derive more reliable cardiac parameters useful to clinicians. This proposed algorithm is introduced in section II, and is tested on SCG signals collected from 25 subjects

II. Methodology

In order to detect and remove the low-quality SCG cycles, our methodology included different stages as are discussed in the following sections.

A. Data collection, Segmentation, Preprocessing, and Labeling

The seismocardiogram (SCG) and electrocardiogram (ECG) of 25 subjects with ischemic heart disease (IHD) were recorded at Burnaby General Hospital under an ethics approval (Age=66.5±8.2 years, Weight=84.3±20.7 kg, Height=170±8.9 cm, 11 females / 14 males). Written consent had been received from every subject. The subjects with IHD have a higher probability of giving low-quality (LQ) SCG cycles compare to the healthy subjects. Therefore, the capability of the proposed approach can be precisely investigated with respect to these LQ cycles.

The SCG data were segmented based on the heartbeats using the R-peaks of ECG signals. The total number of 1697 cycles was obtained and each cycle was preprocessed to have a zero mean and unit variance. After preprocessing, an expert carefully assessed all cycles, and labeled 248 cycles as LQ cycles. The labeled cycles were severely affected by noise, respiration, heart abnormality, and/or other artifacts, so that the expert was not able to distinguish the key points of SCG that were correlated to cardiac activities. The expert was also blind to further processing of the data.

B. Signal quality indices

Different features were extracted from SCG cycles and their corresponding systolic and diastolic phases, which are referred to as signal quality indices (SQIs). The systolic and diastolic phases were obtained using the method described in [9]. The extracted SQIs were as follow.

D.A. Jaffray (ed.), *World Congress on Medical Physics and Biomedical Engineering, June 7-12, 2015, Toronto, Canada*,
IFMBE Proceedings 51, DOI: 10.1007/978-3-319-19387-8_ 247

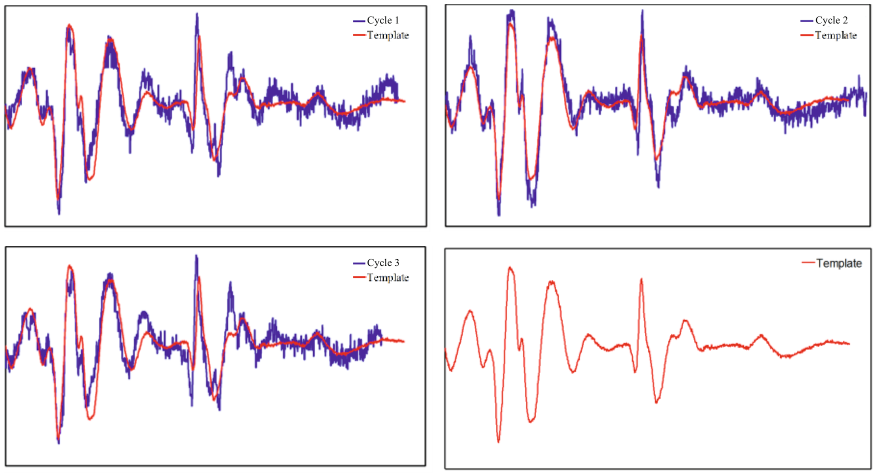

Fig. 1 Three SCG cycles and their computed template.

Template: Template is a representative cycle that has the maximum correlation with each of the cycles. The algorithm that computes template is iterative and is based on the Woody's method [10]. In this algorithm, each cycle is shifted with respect to the average of all cycles until obtaining the maximum correlation between the cycle and the average. The shifted cycles would obtain a new average, and the algorithm can be iterated to reach an average with the desired accuracy that is called the template. Template has been recently employed in quantification of morphological changes among SCG cycles [6]. Fig. 1 shows 3 SCG cycles and their computed template.

In this study, the maximum lag and lead shifts was chosen to be 50 samples, and the number of iterations was selected to be 10. The shift and correlation of each cycle with respect to their template were chosen as the SQIs.

Kurtosis: Kurtosis indicates the peakedness of a probability distribution. Previous research indicated the capability of using Kurtosis as a SQI for ECG signals for identifying muscle artifact, baseline wander, and power-line interference [11]. Therefore, it would be promising to use this SQI for identifying SCG LQ cycles. Kurtosis is computed as the fourth moment of the distribution ($E\,(x-\mu)^4/\sigma^4$); where x, μ, and σ are data, its mean, and its variance, respectively.

Skewness: Skewness indicates the asymmetry of a probability distribution. It has been used as a SQI to assess the quality of ECG signals during arrhythmia [12]. Therefore, it has a high potential to identify LQ cycles of SCG. Skewness can be computed as the third moment of the distribution ($E\,(x-\mu)^3/\sigma^3$).

Power: A signal's power in the time domain, and/or its power spectral density (PSD) in the frequency domain can be significantly affected in the presence of noises and other artifacts. Therefore, these features were selected as SQIs. Moreover, the ratios of systolic power/PSD to diastolic power/PSD were considered as other SQIs, and their performance was evaluated. The signal power (SP) was computed as $\sum_{i=1}^{n} x_i^2/n$; where n is the number of samples in the dataset x. The power spectral density was computed using the Burg method [13].

Number of Minimums/Maximums: In an SCG cycle, there are maximum and minimum points that are correlated to cardiac activities [1]. Fig. 2 shows an SCG cycle with the indicated points, which are corresponded to iso-volumetric moment (IM), aortic valve opening (AO), and aortic valve closure (AC). Identifying these points on SCG and annotating them appropriately is critical, because they can employed to compute hemodynamic parameters such as systolic time intervals [2].

As a result, if such points cannot be identified properly on an SCG cycle, the cycle should be labeled as LQ. In the annotation process of such points, AO is mostly identified based on IM (the first maximum after it), and therefore it is sufficient to find IM and AC [9]. Two appropriate SQIs can be counting the number of local minimums and maximums

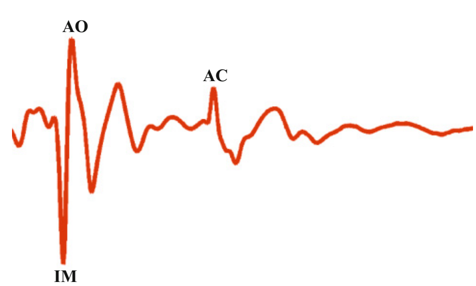

Fig. 2 The maximum and minimum points on an SCG cycle that are corresponded to iso-volumetric moment (IM), aortic valve opening (AO), and aortic valve closure (AC).

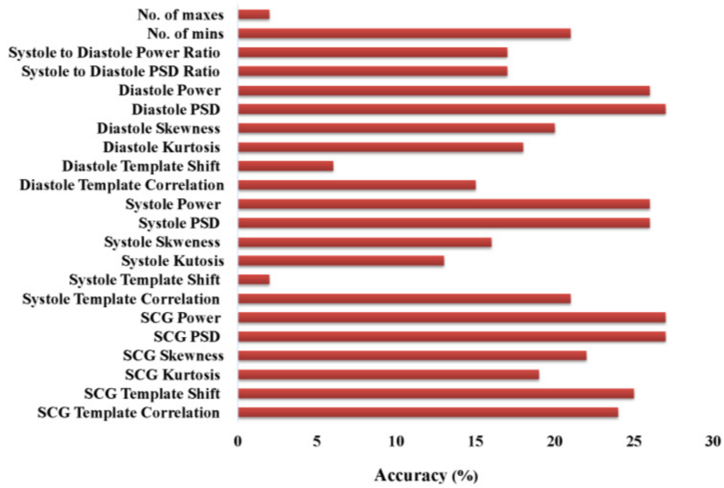

Fig. 3 The accuracy of each SQI

in small windows around IM and AC respectively. Large or zero number of minimum and maximum points implies the difficulty of identifying IM and AC; and therefore it is beneficial to utilize such SQIs. In order to obtain the windows around IM and AC, the method descried in [9] was employed.

C. Detection and removing algorithm

To detect and remove the LQ cycles, the proposed algorithm performed the following procedure on each subject,

(1) The SCG cycles were grouped based on the heartrate (HR). These HR groups were [the unit is beat per minute (BPM)]: (i) HR < 50, (ii) 50 ≤ HR < 60, (iii) 60 ≤ HR < 70, (iv) 70 ≤ HR < 80, (v) 80 ≤ HR < 90, (vi) 90 ≤ HR < 100, (vii) 100 ≤ HR.

(2) For each HR group, the SQIs were computed.

(3) For the data-points of each SQI, the distance between 25th and 75th percentiles was computed (d_1). Then, an interval was set around the median point. The distance from its high and low boundaries to the median was $d_2 = 1.5\ d_1$.

(4) An SCG cycle associated with a data-point outside of the interval, obtained in the previous step, was considered as an outlier.

(5) The "repetition number" (RN) for each outlier was determined by counting the number of SQIs obtaining that outlier.

(6) The outliers were ordered based on their RN in a descending fashion. Starting from the highest RN, the associated SCG cycle was removed. The removal process was continued for the cycles associated to next higher RNs.

(7) The removal process in step (6) stopped, if either there were no outliers, or the total number of remained cycles (among all HR groups) was 10.

The hypothesis in designing the above algorithm was that the set of total outliers would include the set of LQ cycles labeled by the expert. Therefore, removing the outliers would remove the LQ cycles. The algorithm could also mis-detect the high quality cycles as low quality ones. This case was not a major concern, because as it was mentioned earlier, in real practice 10 cycles are sufficient for cardiac assessment.

One advantage of the designed algorithm is that first the outliers are removed that have higher RNs. A higher RN means the outlier is common among more SQIs, which implies its greater probability to be an LQ cycle. As a result, the remained cycles would be less prone to have LQ cycles. Another advantage is that the algorithm determines all its parameters automatically (e.g. median and percentiles), and there is no need for training. This "unsupervised approach" makes the algorithm applicable when no labeled data are available.

III. RESULTS

The proposed algorithm was implemented in the MATLAB platform, and was tested on the SCG signals collected from 25 subjects. The detected outliers had a great overlap with the LQ cycles that were manually labeled by the expert. Among total 1697 SCG cycles, the algorithm could successfully detect 208 out of 248 LQ cycles (84%) when all the features were used. In order to find which feature had the highest impact in the LQ detection, the individual effect of each feature was studied, which is shown in Fig. 3

As Fig. 3 indicates, the "diastole power" had the maximum accuracy (27.4 %) among all SQIs, whereas the "number of maximums" around AC had the minimum accuracy (1.6 %). Generally, the power and PSD of cycles (SCG, systole, and diastole) had higher accuracies comparing to other SQIs. Systole and diastole template shifts as well as the number of maximums had the lowest accuracies. Also, the analysis of these results for kurtosis and skewness revealed that these SQIs performed efficiently for SCG cycles as we anticipated based on their capability as SQIs for ECG.

IV. Conclusions

This paper presented an algorithm to automatically detect and remove low-quality (LQ) cycles for seismocardiogram (SCG). The proposed algorithm extracted 22 signal quality indices (SQIs), computed the outliers, and removed them accordingly. The main advantage of the algorithm was its unsupervised approach, in which there was no need for training by some known labeled data. The focus of the developed algorithm was on detecting the LQ cycles, and not classifying low and high-quality cycles. This goal was considered based the real practice in which 10 cardiac cycles are sufficient for analysis.

The performance of the algorithm was tested on 248 manually labeled LQ cycles, in which 84% of such cycles were detected. The obtained accuracy is promising, and establishes a solid ground for further analysis, and improving the algorithm. As part of our future plans, we would evaluate the algorithm on a larger number of labeled data, provided by different experts. Also, we would modify the algorithm accordingly for better efficiency and higher accuracy.

Acknowledgment

We would like to thank Burnaby General Hospital for providing access to the patients and Heart Force Medical Inc. for providing the data acquisition system. We would also like to thanks Alireza Akhbardeh for his helps.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Zanetti J M, Salerno D (1990) Seismocardiography: A new technique for recording cardiac vibrations: concept, method, and initial observations. J. Cardiovasc. Technol., vol. 9, no. 2, 111–118
2. Zanetti J M, Tavakolian K (2013) Seismocardiography : past, present and future. IEEE Engineering in Medicine and Biol-ogy Society Conference, 7004–7007
3. Castiglioni P, Faini A, Parati G, Di Rienzo M (2007) Wearable seismocardiography. IEEE EMBC conference, 3954–3957
4. Paukkunen M, Parkkila P, Hurnanen T, Pankaala M, Koi-visto T, Nieminen T, Kettunen R, Sepponen R (2015) Beat-by-beat quantification of cardiac cycle events detected from three-dimensional precordial acceleration signals. IEEE J. Biomed. Heal. Informatics (in press)
5. Tavakolian K, Vaseghi A, Kaminska B (2008) Improvement of ballistocardiogram processing by inclusion of respiration information. Physiol. Meas., vol. 29, no. 7, 771–781
6. Zakeri V, Tavakolian K, Arzanpour S, Zanetti J M, Dumont G A (2014) Preliminary results on quantification of seismocardiogram morphological changes, using principal component analysis. IEEE Engineering in Medicine and Biology Conference, 6092–6095
7. Boudoulas H (1990) Systolic time intervals. Eur. Heart J., vol. 11, 93–104
8. Lewis R P, Rittogers S E, Froester W F, Boudoulas H (1977) A critical review of the systolic time intervals. Circulation, vol. 56, no. 2, 146–158
9. Khosrow-khavar H, Tavakolian K, Blaber A P, Zanetti J M, Fazel-rezai R, Menon C (2014) Automatic annotation of seismocardiogram with high frequency precordial accelerations. IEEE J. Biomed. Heal. Informatics (in press)
10. Woody C D (1967) Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals. Med. Biol. Eng., vol. 5, no. 6, 539–554
11. Li Q, Mark R G, Clifford G D (2008) Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman Filter. Physiol. Meas., vol. 29, no. 1, 15–32
12. Behar J, Oster J, Li Q, Clifford G D (2013) ECG signal quality during arrhythmia and its application to false alarm reduction. IEEE Trans. Biomed. Eng., vol. 60, no. 6, 1660–1666
13. Burg J P (1967) Maximum entropy spectral analysis. Proceedings of the 37th Annual International SEG Meeting vol. 37

Corresponding Author: Vahid Zakeri
Institute: Heart Force Medical Inc.
Street: 666 Burrard Street, Suite 500
City: Vancouver
Country: Canada
Email: vahid888@gmail.com